Beutnagel 3-12-9

IN THE CLAIMS:

1. (Presently Amended) A method for generating a signal rich in prosody information comprising the steps of:

inserting in said signal a plurality of phonemes represented by phoneme symbols,

inserting in said signal a duration specification associated with each of said phonemes,

inserting, for at least one of said phonemes, at least two prosody parameter specifications, with each specification of a prosody parameter specifying a target value for said prosody parameter and any selected point in time for reaching said target value, to thereby generate a signal adapted for converting into speech.

2. The method of claim 1 where at least one of said two prosody parameter specifications specifies pitch.

3. The method of claim 1 where at least one of said two prosody parameter specifications specifies energy.

4. The method of claim 1 where a first of said two prosody parameter specifications is a pitch specification and a second of said two prosody parameter specifications is an energy specification.

5. The method of claim 1 where either one of said at least two prosody specifications specifies an energy with a target value corresponding to silence.

Claim **6 - Previously Deleted.**

7. The method of claim 1 where said point in time for reaching target value of a specified prosody parameter of a phoneme from said plurality of phonemes is expressed in terms of time offsets from the beginning of phonemes.

Claims **8 and 9 - Previously Deleted.**

2

10. The method of claim 1 where said signal also includes text specifications.

Claims **6 and 12 - Previously Deleted.**

13. The method of claim 10 where said signal also includes image specifications.

14. The method of claim 1 where said point in time is specified as an offset from beginning of said one of said phonemes.

15. The method of claim 1 where said at least two prosody parameter specifications comprise at least two pitch specifications.

16. The method of claim 1 where said at least two prosody parameter specifications comprise at least two pitch specifications followed by an energy specification.

17. The method of claim 1 where said at least two prosody parameter specifications comprise a plurality of one or more pitch specifications and a plurality of one or more energy specifications.

18. The method of claim 1 where said at least one of said phonemes includes more than two prosody parameter specifications, with each specification of a prosody parameter specifying a target value for said prosody parameter and a point in time for reaching said target value

19. The method of claim 18 where each of at least two of said more than two parameter specifications specifies a pitch target value and a time for reaching said pitch target value.

20. The method of claim 18 where each of at least two of said more than two parameter specifications specifies an energy target value and a time for reaching said energy target value.

21. (Presently Amended) A method for generating a signal rich in prosody information comprising:

a first step for inserting in said signal a plurality of phoneme symbols,

a second step for inserting in said signal a desired duration of each of said phoneme symbols,

a third step for inserting at least one target prosody parameter value within a duration for at least one of said phonemes at an explicitly chosen time offset from the beginning of the duration of said phoneme that is greater than zero and less than the duration of said phoneme.

22. A method of claim 21 where said prosody parameter value is unrestricted at other than said chosen time offset.

23. (Newly Added) The method for creating a signal responsive to a text input the results in a sequence of descriptive elements, including, a TTS sentence ID element, a silence duration specification element, if a silence specification is desired; a gender specification element, if gender specification is desired; an age specification element, if gender specification is desired; a number of text units specification element; and a detail specification the text units, **the improvement comprising the step of**:

including in said detail specification of said text units

- preface information that includes indication of number of phonemes,
- for each phoneme of said phonemes, an indication of number of parameter information tuples, N, and
- N tuples, each tuple specifying a prosody parameter target value and a selectably chosen point in time for reaching said target value.

4

24. (Newly Added) The method of claim 23 where said text units are bytes of text.

25. (Newly Added) The method of claim 23 where said parameter information tuples relate to pitch.

26. (Newly Added) The method of claim 23 where N is an integer greater than 1.

27. (Newly Added) The method of claim 23 where said preface includes a Dur_Enable indicator and when said Dur_Enable indicator is at a predetermined state, said step of including also includes, a phoneme duration value for each phoneme of said phonemes.

28. (Newly Added) The method of claim 23 where said preface includes an F0_Contour_Enable indicator that is set at a predetermined state when said signal includes said N tuples.

29. (Newly Added) The method of claim 23 where said preface includes a Energy_Contour_Enable indicator, and when said Energy_Contour_Enable indicator is at a predetermined state, said step of including also includes, one or more energy value parameters.

30. (Newly Added) The method of claim 29 where said energy value parameters specify energy at beginning, middle, or/and end of phoneme pertaining to said Energy_Contour_Enable indicator.

31. (Newly Added) The method of claim 23 where said preface includes a listing of said phonemes.

32. (Newly Added) A method for generating a signal for a chosen synthesizer that employs text, phoneme, and prosody information input to generate speech, comprising the steps of:

receiving a first number, M, of phonemes specification;

receiving, for at least some phoneme, a second number, N, representing number of parameter tuples to be received for the phoneme;

receiving N parameter tuples, each tuple specifying a parameter target value and a time for reaching said target value;

translating said parameter tuples to form translated prosody information that is suitable for said chosen synthesizer; and

including said translated prosody information in said signal.

33. (Newly Added) The method of claim 32 further comprising:

a step, preceding said step of receiving said second number, M phoneme specifications; and

a step of including in said signal phoneme specification information pertaining to said received M phoneme specifications, which information is compatible with said chosen synthesizer.

34. (Newly Added) The method of claim 32 further comprising the steps of

receiving, following said step of receiving said N parameter tuples, energy information; and

including in said signal a translation of said energy information, which translation is adapted for employment of the translated energy information by said chosen synthesizer.